

審査の結果の要旨

氏名 吉永直樹

語彙化文法は、単語に対する構文的制約と意味的制約を記述した語彙項目を、比較的少量の文法規則を用いて組み合わせることで、文に対し構文構造・意味表現を与える文法枠組である。語彙化文法は、構文や意味構造に関する深い解釈を可能とする一方、それを自然言語処理に用いるためには言語処理技術においていくつかの困難があった。本論文は、自然言語処理に語彙化文法を用いるために、個々の語彙化文法の枠組み間で文法資源及び構文解析技術を共同開発するための方法論と語彙化文法のための辞書資源をコーパスから獲得する方式を新たに開発し、その方法論及び方式の有効性を理論的考察と計算機実験を通して実証している。

本論文は2部からなり、これに序章と結論の章が付けられている。第1部は語彙化文法間の共同開発の方法論に関するもので、このために第1章から4章が当てられている。第2部は、第5章から7章からなり、コーパスからの辞書資源の獲得方式について論じている。

第1章は、語彙化文法に関する形式的概念についてであり、その主要なものとして **Lexicalized Tree Adjoining Grammar (LTAG)** と **Head-Driven Phrase Structure Grammar (HPSG)** について利用可能なリソースも含め解説している。

第2章は、LTAG の枠組みを HPSG の枠組へ変換するための文法変換アルゴリズムを開発し、強い意味で等価な文法に変換できることを証明している。そして、この文法変換を個々の枠組の表層的な違いを捨象する手段として用い、語彙化文法に対する汎的な言語処理技術を開発する方法論について述べている。これにより、等価な文法を仲介して構文解析器、曖昧性解消モジュール、文法開発環境などの言語処理技術の共有を可能にしている。

第3章では、LTAG から HPSG スタイルの文法への文法変換を実装し、ペンシルバニア大学で開発された大規模 LTAG 英語文法を HPSG へ変換する計算機実験を行っている。この結果得られた等価であることが保証されたな文法を用いて、動的計画法と CFG フィルタリングとよばれる構文解析手法について LTAG と HPSG とで別々に設計された構文解析器を比較・解析している。その結果、構文解析法の開発について、文法枠組の違いを越えて共同的に開発することが可能になることを実証している。第4章は、本論文の研究に関連した研究について言及している。

第5章は、下位範疇化フレーム(**Subcategorization Frame, SCF**)を導入し、語彙化文法の中でそれを取り扱う方法について述べている。そして、生コーパスから語彙化文法のための SCF を自動的に抽出する方法について論じている。

第6章は、生コーパスから抽出されたノイズを含む SCF のフィルタリングを行うための方法について論じている。その方法は、増強対象になっている辞書中の単語と生コーパス

から抽出した単語とを、それらが取る SCF の組み合わせに従ってクラスタリングし、得られた結果を用いて SCF のフィルタリングを行うもので、計算機実験では、大規模語彙化文法の辞書に対し、辞書に含まれる情報を用いることで未知語のより正確な SCF を選別することに成功している。

第7章は、増強対象となっている辞書が、単語と SCF の共起確率を含む場合に、それらの確率を対象ドメインに対して適切な確率にスムージングする手法を提案している。

このようにして5章から7章は、既存の文法リソースと統合するために十分な正確さ備えた語彙化文法の文法リソースを、生コーパスから獲得する方法について論じ、その有効性を計算機実験を通して証明している。

なお、本論文の内容は、辻井潤一、鳥澤健太郎、宮尾祐介、建石由佳氏との共同研究であるが、論文提出者が主体となって分析及び検証を行ってもので、論文提出者の寄与が十分であると判断する。

よって本論分は博士（情報理工学）の学位申請論文として合格認められる。